The EMBO Journal vol.8 no.5 pp.1497 - 1505, 1989

The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes

Denis Duboule[1] and Pascal Dollé

Laboratoire de génétique moléculaire des eucaryotes du CNRS, U.184 de biologie et de génie génétique de l'INSERM, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg cédex, France

[1]Present address: EMBL, Meyerhofstraße 1, Postfach 10.22 09, D-6900 Heidelberg, FRG

Communicated by P.Chambon

This paper reports the cloning of the fourth major murine homeogene complex, HOX-5. The partial characterization of this gene cluster revealed the presence of two novel genes (*Hox-5.2, Hox-5.3*) located at the 5′ extremity of this complex. *In situ* hybridization experiments showed that these two genes are transcribed in very posterior domains during embryonic and foetal development. We also show that *Hox-1.6*, the gene located at the 3′ most position in the HOX-1 complex, has a very anterior expression boundary during early development. These results clearly support the recently proposed hypothesis that the expression of murine Antp-like homeobox-containing genes along the antero-posterior developing body axis follows a positional hierarchy which reflects their respective physical positions within the HOX clusters, similar to that which is found for the *Drosophila* homeotic genes. Such a structural and functional organization is likely conserved in most vertebrates. Moreover, on the basis of sequence comparisons, we propose that the ordering of homeobox-containing genes within clusters has been conserved between *Drosophila* and the house mouse. Thus, very different body plans might be achieved, both in insects and vertebrates, by evolutionarily conserved gene networks possibly displaying similar regulatory interactions.

Key words: homeobox/mouse embryo/positional cues/evolution/development

Introduction

Antennapedia-like homeobox-containing genes have been isolated from many different invertebrate and vertebrate species (see McGinnis, 1985; Holland and Hogan, 1986; Gehring, 1987). In *Drosophila*, the homeobox-containing homeotic genes are essentially dispersed in two clusters: the Bithorax (BX-C; Lewis, 1978; Sanchez-Herrero *et al.*, 1985) and Antennapedia (ANT-C; Kaufmann *et al.*, 1980) complexes, whose internal structural organization has been shown to reflect the order in which these genes are expressed along the antero-posterior body axis during development. Thus, the relative position of each of these genes within the complexes correlates with their expression domains along the antero-posterior axis and, as a result, the structures that are specified by these different domains (for review and refs see Harding *et al.*, 1985; Akam, 1987; Scott and Carrol, 1987). In the house mouse, >20 Antp-like homeobox-containing genes have been reported so far to lie on four major complexes: HOX-1, -2, -3 and -5, located on chromosomes 6, 11, 15 and 2 respectively (see Figure 3 for refs). Based on protein sequence similarity, these genes can be grouped into subfamilies which are represented once in each cluster. Since each subfamily member is found in the same relative position within each cluster (Hart *et al.*, 1987; Duboule *et al.*, 1989; Graham *et al.*, 1988), the HOX-1, -2, -3 and -5 complexes are likely to be the result of large-scale duplication events during evolution.

Accumulating evidence indicates that murine *Hox* genes (homeogenes) play a crucial role in embryonic and foetal development. Because they are expressed during ontogeny in the same types of structure (central and peripheral nervous system, somitic and non-somitic mesodermic derivatives, limb buds) but in different overlapping antero-posterior domains (Awgulewitsch *et al.*, 1986; Utset *et al.*, 1987; Gaunt, 1987, 1988; Gaunt *et al.*, 1986, 1988; Dony and Gruss, 1987; Holland and Hogan, 1988a) it was suggested that vertebrate homeoproteins might serve as positional cues along the rostro-caudal axis of the developing animal. It was recently further proposed (Gaunt *et al.*, 1988) that, as in *Drosophila*, the ordering of the HOX genes along the various complexes may also reflect the antero-posterior distribution of their expression domains and that gene members of the same subfamily might therefore display coincident anterior expression boundaries (Gaunt *et al.*, 1988). In this paper, we present the structure of the fourth complex, HOX-5, and show that it contains at least two other members in addition to the previously reported *Hox-5.1* (Featherstone *et al.*, 1988). These two newly described members, *Hox-5.2* and *Hox-5.3*, are located at the most 5′ positions so far described for any HOX cluster. *In situ* hybridization studies reveal that they are expressed in the posterior-most regions of the foetus. We also show that the gene so far described as lying at the most 3′ extremity of any complex (*Hox-1.6*, Baron *et al.*, 1987) displays a very anterior expression border. Based on these observations and on significant similarities to sequences and relative domains of expression of homeotic genes of *Drosophila*, we suggest that both the structural and functional organization of the homeobox-containing gene families has been conserved between insects and vertebrates.

Results

Molecular cloning of the HOX-5 complex

We recently reported the isolation and description of the homeogene *Hox-5.1* on mouse chromosome 2 (Featherstone *et al.*, 1988). *Hox-5.1* forms a subfamily with the closely related *Hox-1.4* and *Hox-2.6* homeogenes, in the HOX-1 and HOX-2 complexes respectively (B.Galliot and D. Duboule, unpublished results; Graham *et al.*, 1988). In order to see if this novel HOX-5 complex would contain other homeogenes, possibly similar to those of the HOX-1, -2 and -3

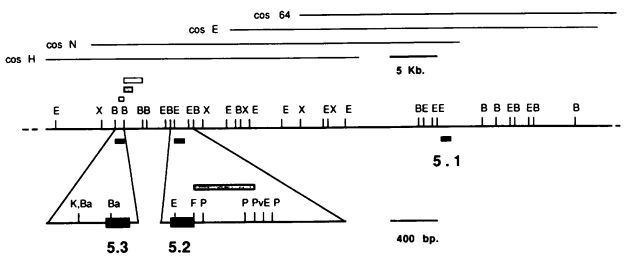

Fig. 1. Partial molecular map of the HOX-5 complex. About 60 kb of DNA are shown as defined by the four overlapping cosmid clones shown at the top. The black rectangles represent homeoboxes. The bottom line shows enlargements of two subclones containing the 5.3 and 5.2 homeobox and flanking sequences. The open or shaded rectangles are different probes used in *in situ* hybridization experiments (see text). The three genes have the same orientation with respect to transcription: 5′ to 3′ from left to right. E, *Eco*RI; X, *Xho*I; B, *Bam*HI; K, *Kpn*I; Ba, *Ban*I; F, *Fok*I; P, *Pst*I; Pv, *Pvu*II.

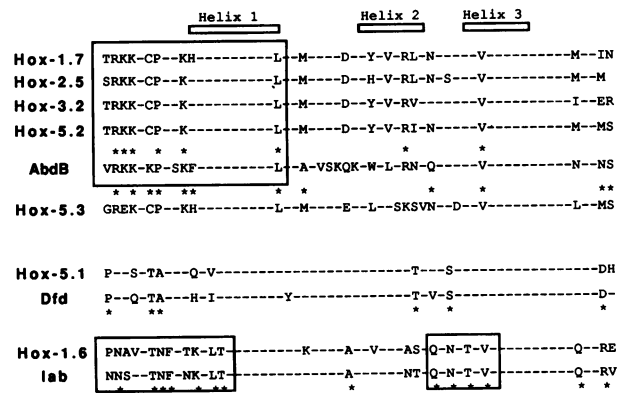

Fig. 2. One-letter code protein sequence of the *Hox-5.2* and *Hox-5.3* homeo-domains and their similarities to other murine homeo-domains. Only those amino acids which differ from the Antennapedia 'reference' sequence are shown (see Hart *et al.*, 1987; Duboule *et al.*, 1989). The top four lines show the *Hox-5.2* sequence aligned with the three other sequences of the genes belonging to the same (*Hox-1.7*-like) subfamily (on a vertical line in Figure 3). The boxed domain indicates a very divergent N-terminal part sharing high similarity with the *AbdB* gene (the stars between *Hox-5.2* and *AbdB* indicate identical amino acid substitutions). The *Hox-5.3* sequence is given below with, on the top, stars indicating amino acids found in the *Hox-1.7*-like subfamily. The bottom part of the figure shows two sequence comparisons between *Drosophila* and murine homeo-domains (see the text and Figure 6). Sequence data are from: *Hox-1.7*, Rubin *et al.* (1987); *Hox-2.5*, R.Krumlauf, personal communication; *Hox-3.2*, Breier *et al.* (1988); *AbdB* (Iab7), Regulski *et al.* (1985); *Hox-5.1*, Featherstone *et al.* (1988); *Dfd*, Regulski *et al.* (1987); *Hox-1.6*, Baron *et al.* (1987); *lab*, Hoey *et al.* (1986) and Mlodzick and Gehring (1988).

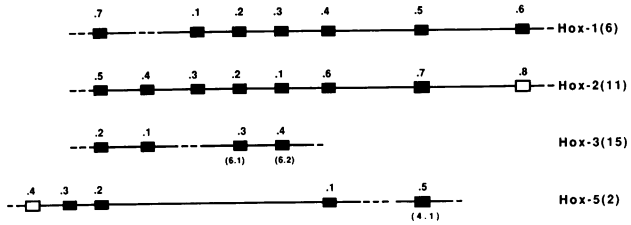

Fig. 3. Schematic and partial representation of the murine HOX family. The four HOX complexes are aligned according to the similarities between genes that are members of the same subfamily (see Duboule *et al.*, 1989). Thus, genes located along the same vertical line show high degrees of similarity. The filled boxes represent homeobox genes which have been sequenced, whereas open boxes are putative candidates not yet sequenced. Numbers within parentheses refer to chromosomes. The *Hox-6.1* and *Hox-6.2* genes have been reassigned to chromosome 15 and are therefore tentatively considered as being linked to the HOX-3 complex (see Gaunt *et al.*, 1988). Similarly, the *Hox-4.1* gene (Lonai *et al.*, 1987) may be a member of the HOX-5 complex. In both cases, physical linkages have not yet been reported as in the case of *Hox-1.7* and the HOX-1 complex. Data are from: the HOX-1 complex (Colberg-Poley *et al.*, 1985; Duboule *et al.*, 1986; Baron *et al.*, 1987; Rubin *et al.*, 1987); the HOX-2 complex (Hart *et al.*, 1985, 1987; reviewed in Graham *et al.*, 1988; and R.Krumlauf, personal communication); the HOX-3 complex (Awgulewitsch *et al.*, 1986; Breier *et al.*, 1988; Sharpe *et al.*, 1988); the HOX-5 complex (Featherstone *et al.*, 1988; this work).

complexes, four overlapping cosmid clones from the HOX-5 locus were screened by Southern blot analysis with a battery of homeobox-containing probes. Unlike the HOX-1 or HOX-2 complexes, no homeoboxes were detected in the 30 kb upstream from *Hox-5.1* (Figures 1 and 3). Instead, two cross-hybridizing sequences were found, 31 and 37 kb upstream of *Hox-5.1* (Figure 1), and were therefore called *Hox-5.2* and *Hox-5.3*. The putative protein sequence of the *Hox-5.2* homeobox and flanking regions revealed a very high degree of sequence similarity with *Hox-1.7*, *Hox-2.5* and *Hox-3.2* (Figure 2). *Hox-5.2* is thus the fourth member of the *Hox-1.7* subfamily. Interestingly, the distance between *Hox-5.1* and *Hox-5.2* is comparable to that between their HOX-2 homologues, *Hox-2.6* and *Hox-2.5* (Graham *et al.*, 1988) even though the homologues to the four *Hox-2* genes found in between do not seem to be present in the HOX-5 complex (Figure 3). The *Hox-5.2* homeo-domain sequence is characterized by two domains of high divergence when compared to the Antp reference sequence, in the putative helix 2 domain and at the N terminus. In this last case, six out of the eight amino acids exchanged are identical to those present in the *Drosophila* AbdB homeo-domain (Figure 2), whereas such changes are not found in any other Antp-like homeo-domains. Other such significant sequence similarities between *Drosophila* homeotic genes and mouse *Hox* genes have been previously reported (*Hox-1.6* versus *lab*; Duboule *et al.*, 1989; *Hox-5.1* versus *Dfd*; Featherstone *et al.*, 1988) and are included in this figure for discussion (see below). The *Hox-5.3* homeo-domain sequence also resembles *Hox-1.7*, to some extent, though a very different stretch of amino acids in the helix 2 region suggests it to be the first member of a new subfamily located at the 5′ most position of any HOX complex (Figure 3). Alternatively, *Hox-5.3* and *Hox-5.2* may have arisen through a late duplication event, specific to the HOX-5 complex. The presence or absence of *Hox-5.3*-related sequences in the other HOX complexes will answer this question.

Expression of the Hox-5.2 and Hox-5.3 genes along the rostro-caudal axis of developing foetuses

We have used the *Hox-5.2* and *Hox-5.3* genes to test the recently proposed hypothesis that the more 5′ a HOX subfamily is located in a complex, the more posterior will be their expression domains during development (Gaunt *et al.*, 1988). Thus, the extreme 5′ position of the *5.3–5.2* tandem in the HOX network (Figure 3) should be reflected, according to this model, by very posterior expression domains during foetal development. As seen in Figure 4, this is in fact the case. *Hox-5.2* is strongly expressed in 12.5 day old foetuses in both spinal cord and pre-vertebral column with an anterior boundary of expression lying at the level of the first lumbar pre-vertebrae (Figure 4B and F), 9–10 metameric units more posterior than *Hox-3.1*, the gene

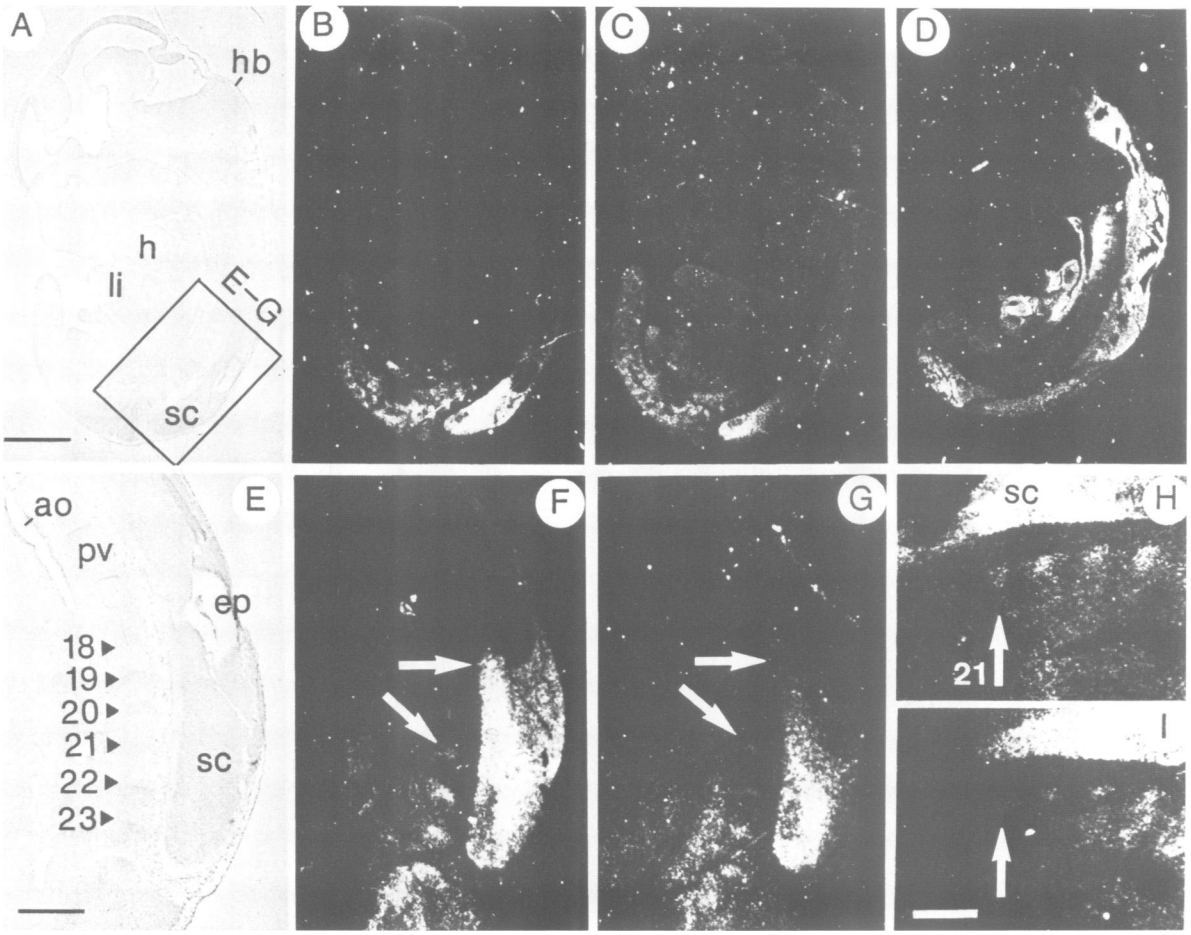

Fig. 4. The antero-posterior boundaries of expression of the *Hox-5.2* and *Hox-5.3* genes. *In situ* hybridization on a 12.5 day old mouse foetus. (**A**,**B**) Sagittal section hybridized with the *Hox-5.2* probe and seen in bright field (A) or dark field (B) illumination. The antero-posterior axis of the foetus is slightly curved so that the plane of section is medial at the tail and lumbar levels but more lateral in the cervical region. Bar: 1.0 mm. (**C**) The neighbouring section hybridized to the *Hox-5.3* probe. (**D**) A comparable sagittal section at the same age hybridized with a *Hox-1.4* probe (same as in Gaunt *et al.*, 1988) is used as a positive and 'anterior' control. Adjacent sections were systematically hybridized to sense RNA probes as negative controls and no specific signals were obtained (not shown). The difference between the 'posterior' expression of 5.2 and 5.3 and the 'anterior' expression of 1.4 is evident. (**E**) Enlargement of the rectangle in (A) with the indication of pre-vertebrae 18–23. Bar: 0.4 mm. (**F**) and (**G**) very clearly show the slight, but reproducible difference between *Hox-5.2* and *Hox-5.3* expression boundaries respectively. (**H**,**I**) Another example of this precise shift between 5.2 (H) and 5.3 (I). Bar: 0.3 mm. Identical levels in (F) and (G) as well as in (H) and (I) are indicated with white arrows. These arrows show the anterior limits of *Hox-5.2* expression in both spinal cord and pre-vertebrae (F,G) or pre-vertebrae only (H,I). In these last two panels, the orientation is antero-posterior from left to right. hb, hindbrain; h, heart; l, liver; sc, spinal cord; ao, aorta; pv, pre-vertebrae; ep, ependymal lumen. pv 21 is the first lumbar pre-vertebrae.

located immediately 3′ in the aligned complexes (Figure 3). The boundary within the central nervous system (CNS) is also very posterior and found at the level of the 10th thoracic pre-vertebrae (Figure 4). As expected, the *Hox-5.3* gene, located upstream from *Hox-5.2*, shows slightly more posterior expression boundaries. The anterior boundary in the pre-vertebral posterior column is consistently found one to two pre-vertebrae posterior to that of *Hox-5.2* (Figure 4C and G), whereas a larger difference is observed in the CNS corresponding to the width of ~2–3 metameric units (compare Figure 4B to C, F to G and H to I). These two genes may therefore represent very posterior determinants, since both of them start to be expressed in the pre-vertebral column at the lumbar level (see also Dollé and Duboule, 1989).

Expression of the Hox-1.6 gene in early developing foetuses

Having verified that the genes located the most 5′ are the more posteriorly expressed during foetal development, we looked at the expression of the *Hox-1.6* gene (Baron *et al.*, 1987). *Hox-1.6* has been reported to be located at the 3′ extremity of the HOX-1 locus (Duboule *et al.*, 1986), and should therefore be an ideal candidate for displaying a very anterior domain of expression. *In situ* hybridization experiments using a *Hox-1.6* probe failed to reveal any specific signal at day 12.5 post-coitum (p.c.), thereby confirming the unusually low level of *Hox-1.6* transcripts at this stage of development (Baron *et al.*, 1987). However, *in situ* hybridization on younger embryos or foetuses revealed the specific pattern of expression shown in Figure 5. At day 8–8.5 p.c., at a stage where the embryo has not yet accomplished its rotation in order to take its final position, *Hox-1.6* transcripts are clearly found in the embryonic mesoderm, anterior to the first somite, up to the cephalic region, at the level of the foregut and developing heart. The anterior boundary of expression in cephalic mesoderm is roughly located at the level of the distal part of the foregut pocket (Figure 5B–G). The *Hox-1.6* expression domain, at this stage, appears therefore to be more anteriorly located

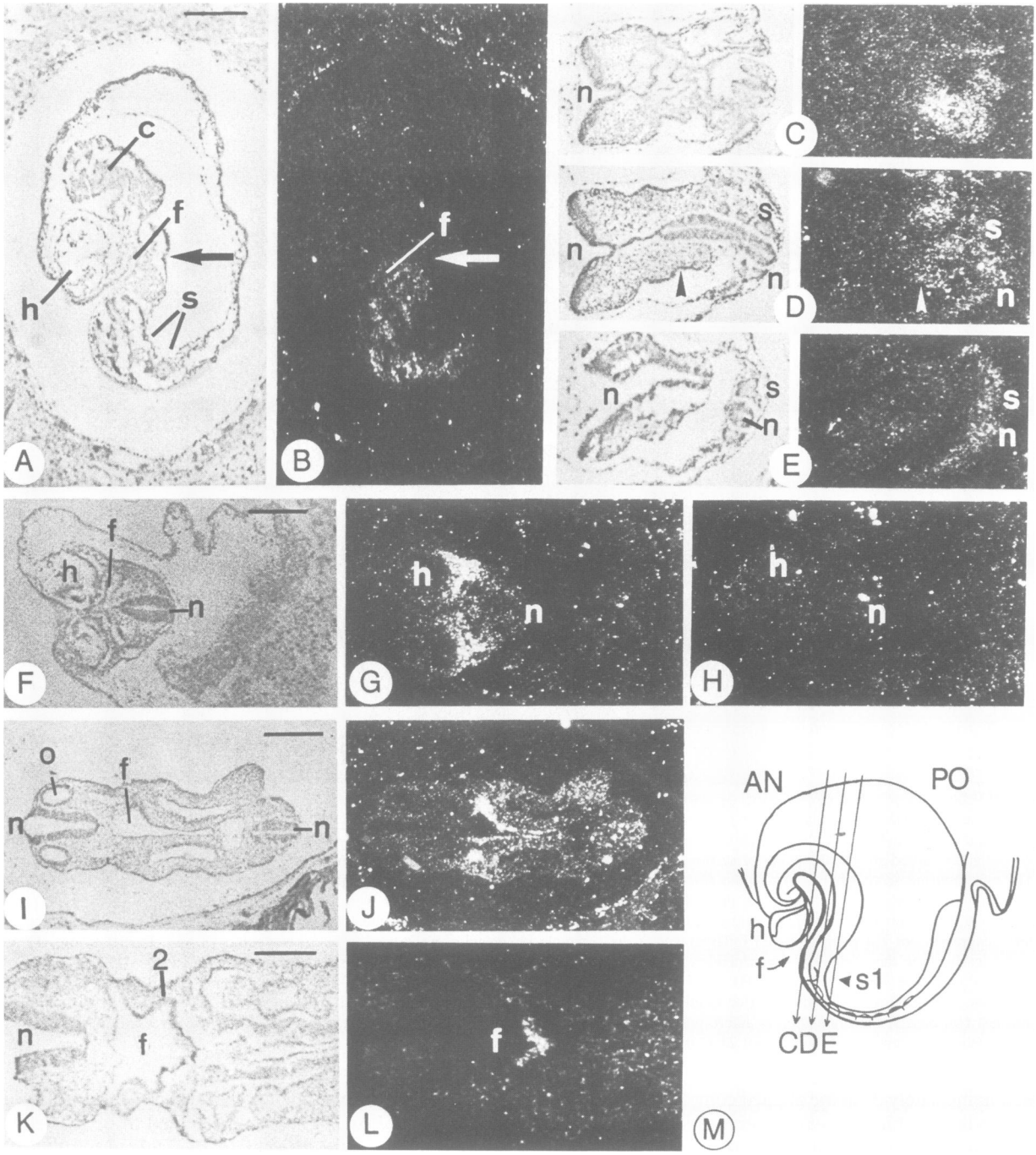

Fig. 5. The *Hox-1.6* expression pattern during early development. (**A**,**B**) Sagittal section of a 8.5 day old embryo (8−10 somites; beginning of rotation) hybridized with a *Hox-1.6* antisense RNA probe and viewed under bright field (A) or dark field (B) illumination. The section is slightly skewed so that anterior and posterior parts are parasagittal. Bar: 0.2 mm. (**C**,**D**,**E**) Serial sections, in a frontal plane, of the anterior part of a young 8.5 day old embryo hybridized to the same probe. The three planes of section are diagrammed in the scheme under (M). Panels A−E illustrate the antero-posterior boundary of *Hox-1.6* expression which lies in a very anterior position (see text). (**F**,**G**,**H**) Section of a 8.5 to 9 day embryo (after rotation) hybridized to either the *Hox-1.6* (G) or the *Hox-5.2* 'posterior' (H) probes. The plane of section is transverse in the heart and foregut regions and a strong signal is found in the epithelium of the foregut as well as in surrounding mesoderm, when the *Hox-1.6* probe is used (G). As expected, no signal is seen with the 'posterior' *Hox-5.2* probe (H). Bar: 0.2 mm. (**I**,**J**) Section of a 9.5 day old foetus, crossing the otic vesicles and the foregut and hybridized to the *Hox-1.6* probe. The signal has a weak but clear distribution and is restricted so that the most anterior part of the section (left) is negative. It is particularly intense in epithelial cells in a posterior part of the foregut. Bar: 0.3 mm. (**K**,**L**) Enlargement of the pharyngeal region from a nearby section of the same embryo as in (I,J) hybridized to the *Hox-1.6* probe. In the foregut region, the signal is found immediately posterior to the second branchial pouch, therefore likely at the level of the third branchial arch or pouch. Bar: 0.2 mm. c, cephalic mesoderm; h, heart; s, somites; f, foregut; n, neuroectoderm; o, otic vesicle; AN, anterior; PO, posterior; number 2 in panel (K) indicates the second branchial pouch. The arrows in (A,B) show the anterior boundary of *Hox-1.6* expression. The arrowhead in (D) shows a proneuromeric constriction (between rhombomeres A and B), which seems to correlate with this anterior boundary. The scheme (**M**) showing the plans of section used in (C−E) was redrawn from Theiler (1972).

than that of *Hox-1.5* (see Gaunt, 1987), the most anteriorly expressed gene, to date, within the HOX complexes (Gaunt *et al.*, 1986; Gaunt, 1987, 1988) located just upstream of *Hox-1.6* (Baron *et al.*, 1987). Transcripts encoded by the *Hox-1.6* gene are also found in the neurectoderm, with an anterior limit which seems to correspond to a neuromeric constriction within the future myelencephalon (Figure 5D). This separation between proneuromeres may correspond to the depression between the so-called rhombomeres A and B, this latter being located at the level of the otic placode (Theiler, 1972). It clearly appears, therefore, that *Hox-1.6*, as predicted from the model, is expressed at a very anterior position. In addition, at both 8.5 and 8 days, a weak signal is regularly found in the most posterior parts of the developing animals (see Figure 5E or J as examples) as for other genes (Holland and Hogan, 1988a).

At day 9 p.c., an accumulation of *Hox-1.6* transcripts is found in an anterior part of the developing foregut and is restricted to an epithelial cell type (Figure 5I–L). Serial sections show that the expression in the foregut epithelium is found at the level of the second to third branchial bars. The anterior boundary of expressing cells probably lies just posterior to the second branchial pouch (Figure 5L). Thus, *Hox-1.6* is expressed in structures which will participate in the formation of the head or upper cervical structures (see Discussion). This very restricted signal is no longer detected at day 10.5 p.c., whereas all other murine homeogenes so far studied are well transcribed (see Holland and Hogan, 1988a; Gaunt *et al.*, 1988 for review).

Discussion

This paper reports the further characterization of the fourth murine HOX gene complex, HOX-5. In addition to the *Hox-5.1* gene (Featherstone *et al.*, 1988), we show that two other homeogenes, *Hox-5.2* and *Hox-5.3*, are located at an extreme upstream position relative to genes in other clusters. The sequence of the *Hox-5.2* homeobox and flanking regions (not shown), and its positions relative to *Hox-5.1*, support the recent proposal that the four HOX gene complexes were generated by large-scale duplication events in the course of evolution (Hart *et al.*, 1987; Duboule *et al.*, 1989). The *Hox-5.2* homeobox is almost identical, at the protein level, to the sequences of the other members of the *Hox-1.7*-like subfamily, whereas *Hox-5.1* belongs to the *Hox-1.4*-like subfamily (Featherstone *et al.*, 1988). Due to this evolutionary link, one would have expected to find, between *Hox-5.1* and *Hox-5.2*, additional *Hox* genes corresponding to the different subfamilies found at these positions in the other complexes (Graham *et al.*, 1988; Dollé and Duboule, 1983) but none have been detected so far even though the appropriate probes were used at very low stringency. Interestingly, however, the distance between *Hox-5.1* and *Hox-5.2* is conserved with respect to subfamily members in the HOX-2 complex (*Hox-2.6* and *Hox-2.5* respectively; see Dollé and Duboule, 1989). The upstream-located *Hox-5.3* homeobox probably represents the first member of a new subfamily whose members may all share the striking sequence difference observed in the putative α-Helix-2 (Otting *et al.*, 1988) when compared to the *Hox-5.2* homeobox. Such differences could affect the DNA binding specificities (qualitatively and quantitatively) of the respective homeoproteins.

A molecular representation of the rostro-caudal axis

Using the *in situ* hybridization technique, we studied the developmental expression profiles of both the *Hox-5.2* and *Hox-5.3* genes as well as of the *Hox-1.6* gene, this last one being located at the downstream extremity of the aligned HOX complexes (Baron *et al.*, 1987). The *Hox-5.2* and *Hox-5.3* expression profiles corroborate a previously proposed model for these proteins acting as positional cues during development or in the establishment of the body plan (for refs, see Holland and Hogan, 1988a). *Hox-5.2* and *Hox-5.3* are both expressed in the spinal cord, as well as in structures derived from either the somitic (pre-vertebrae), intermediate (foetal uro-genital tract, not shown) or lateral plate (intestinal mesoderm, not shown) mesoderm. Though their domains of expression overlap with those of other homeogenes (see, for example, Gaunt *et al.*, 1988), transcripts from both genes are detected much more posteriorly. These similar, but different (1–2 metameric units), anterior expression boundaries along the antero-posterior axis correlate well with the very few organs which show *Hox-5.2* or *Hox-5.3* expression. The very posterior restriction of the *Hox-5.2* and *Hox-5.3* expression domains and their extreme 5′ positions with respect to all the HOX complexes are in excellent agreement with the proposed hypothesis (Gaunt *et al.*, 1988) that the mouse genome contains a direct molecular representation of the antero-posterior body axis. These results are supported by similar studies on the HOX-2 complex (Graham *et al.*, 1989).

This point is further reinforced by the study of the *Hox-1.6* expression domains. We detected *Hox-1.6* transcripts in embryos at day 8 in both neurectoderm and mesoderm until a very anterior region, beyond the first somite. *Hox-1.6* is also expressed in more posterior somitic mesoderm and neural tube. It seems that *Hox-1.6* transcripts extend in both neurectoderm and mesoderm clearly more anteriorly than those of both *Hox-1.4* and *Hox-1.5*, the genes located just upstream of *Hox-1.6* on the HOX-1 complex (Baron *et al.*, 1987; Fainsod *et al.*, 1987; Gaunt, 1988). However, a direct comparison of both *Hox-1.5* and *Hox-1.6* expression domains is not yet available. It seems that the sharp anterior *Hox-1.6* expression boundary coincides with the constriction separating rhombomeres A and B, though a very precise localization of the boundary is made difficult by the low level of transcripts. This expression pattern thus supports a role for homeogenes in establishing broad antero-posterior domains after the onset of gastrulation (Gaunt, 1988; Holland and Hogan, 1988b). The restricted expression of this gene in the foregut epithelium of day 9 foetuses might illustrate a more specific function of *Hox* genes in morphogenesis. In fact, *Hox-1.6* transcripts are detected with a high intensity in the epithelial layer of the foregut pocket, posteriorly to the second branchial pouch. In this case, one might speculate that *Hox-1.6* is involved in the formation of some cephalic structures (those derived from or around the third visceral arch) such as the hyoid or pharyngeal structures. Interestingly, the more posteriorly located *Hox-3.1* gene is expressed in hindgut endoderm at a comparable age (LeMouellic *et al.*, 1988). By days 10–11, *Hox-1.6* transcripts are below our level of detection though other homeogenes are usually expressed at high levels (Gaunt *et al.*, 1988; Holland and Hogan, 1988a). Such temporal variations were previously reported concerning the onset of *Hox-3.1* and *Hox-1.5* expression. In this case, *Hox-1.5*

transcripts accumulate slightly before those of *Hox-3.1* (Gaunt, 1988). Therefore, the order of the genes along the clusters may not only coincide with the positions of their expression boundaries but also with their timing. Experiments using the *Hox-1.6* probe on very early embryos will address this question.

The structural and functional organization of the HOX genes network is conserved among vertebrates

Homeogenes are found clustered in many other vertebrates such as human, chicken and *Xenopus*. In human, Boncinelli and colleagues have reported an overall structural organization very similar to that found in rodents (Boncinelli *et al.*, 1988; Simeone *et al.*, 1988). It is thus very likely that in both species *Hox* genes function in identical networks. In birds, *Hox* genes are also clustered (A.Kuroiwa, personal communication). In at least one case, clustered genes can be aligned with murine genes, on the basis of high sequence similarities. In this last instance, a chicken gene member of the *Hox-1.7*-like subfamily (which contains *Hox-5.2*) is expressed in a posterior domain (A.Kuroiwa, personal communication). In the African toad *Xenopus*, homeogenes are also clustered (see, for example, Harvey *et al.*, 1986). The paired *Xhox1A* and *Xhox1B* genes can be aligned with the murine *Hox-1.4* and *Hox-1.3* genes and their subfamilies. Although direct comparisons of the domains of expression of these genes in *Xenopus* are not available, it is clear that they have different antero-posterior expression boundaries. Thus, the *XlHbox1* or *Xhox1A* genes (Carrasco *et al.*, 1984; Harvey *et al.*, 1986; Oliver *et al.*, 1988) are expressed more anteriorly than the *XlHbox6* gene, recently reported as a marker for posterior neural differentiation (Sharpe *et al.*, 1987). This last gene, expressed in the posterior part of the neural tube, is certainly a member of the *Hox-1.7*-like subfamily since it harbours the amino acid stretch boxed in Figure 2 (see also Breier *et al.*, 1988). It therefore belongs to the same class of 'posterior determinants' exemplified by *Hox-5.2* in this paper. The similarity of their expression patterns (both in the posterior neural tube; Sharpe *et al.*, 1987 and this work) reflects both their structural relatedness and probable positions in respective clusters. Thus, the concept of structurally and functionally defined homeogene subfamilies should be extrapolated to all vertebrates.

Relationship with the Drosophila homeotic gene complexes

Sequence comparisons between murine homeogenes and *Drosophila* homeotic genes reveal, in some cases, striking similarities. Thus, we recently reported that the mouse *Hox-5.1* protein sequence shows very significant conservation with the *Drosophila Deformed* (*Dfd*) homeotic gene (Duboule *et al.*, 1989; Featherstone *et al.*, 1988; see also Regulski *et al.*, 1987). We also reported (Duboule *et al.*, 1989) that the very specific and divergent homeobox found in the mouse *Hox-1.6* gene (Baron *et al.*, 1987) is practically identical to that present in the *Drosophila* homeotic gene *labial* (*lab*; Mlodzick *et al.*, 1988; Figure 2) formerly called *F90-2* (Hoey *et al.*, 1986). Here again the similarities extend outside the homeobox sequences (Baron *et al.*, 1987; Hoey *et al.*, 1986; Mlodzick *et al.*, 1988; and our unpublished observations). Both *Dfd* and *lab* are located in the ANT-C of *Drosophila* and are involved in the specification of anterior structures. In fact, *lab* is the most proximally located gene so far isolated from the ANT-C (Mlodzick *et al.*, 1988) and displays, as expected (see Harding *et al.*, 1985), a domain of expression slightly more anterior than that of *Dfd* (Chadwick and McGinnis, 1987; Martinez-Arias *et al.*, 1987). After gastrulation, *lab* is expressed in the hypopharyngeal organ and anterior parts of the mandibular lobe. At earlier times, *lab* is found expressed in the presumptive anlagen for these organs, just anterior to the cephalic furrow, which is the region of *Dfd* expression at a comparable developmental stage (Chadwick and McGinnis, 1987; Mlodzick *et al.*, 1988). There is therefore a striking correlation between the anteriority of the *lab* and *Hox-1.6* expression domains and an extreme proximal or 3′ location within their respective clusters. Thus, even though the posterior domain of *lab* expression (Hoey *et al.*, 1986; Mlodzick *et al.*, 1988) cannot be compared with any feature of the *Hox-1.6* expression, there is a clear conservation between *Drosophila* and murine homeobox-containing genes, not only in their structure but also in the functional organization of the complexes (Figure 6). This striking parallel between the proximal part of the *Drosophila* ANT-C (*Dfd*; *lab*) and the 3′ region of the HOX-1 cluster (*Hox-1.4*; *Hox-1.6*) can be further extended since *Antp*, a more distally located gene of ANT-C, shows a homeobox sequence virtually identical to those of the *Hox-1.1* and *Hox-1.2* subfamilies (McGinnis *et al.*, 1984; Hart *et al.*, 1987). As suggested in Figure 6, the *Hox-1.5* homeobox sequence may be expected to show some specific similarity with that of *Drosophila* homeotic gene *proboscipedia* (*pb*; Pultz *et al.*, 1988) since this latter gene is located between *Dfd* and *lab*. More surprising is the presence in the homeobox of the *Drosophila* gene *AbdB* (Regulski *et al.*, 1985) of a very specific block of amino acids, selectively found within the homeoboxes of the vertebrate genes belonging to the *Hox-1.7*-like subfamily (see Figure 2 for refs, and Scott *et al.*, 1988 for a detailed review of structural comparisons of homeobox genes). The *AbdB* gene is part of the Bithorax complex (Lewis, 1978; Sanchez-Herrero *et al.*, 1985) and therefore is involved in the specification of more posteriorly located structures (posterior to the second thoracic segment; Lewis, 1978; see Harding *et al.*, 1985 for refs). We show in this study that *Hox-5.2*, a member of the *Hox-1.7*-like subfamily, is also expressed in very posterior domains and, as such, may specify a 'posterior' position cue. We therefore propose a tentative alignment of both *Drosophila* and vertebrate complexes with the respective anterior expression boundaries along the developing CNSs of the fly and the mouse (Figure 6). In this scheme, *Drosophila* genes from the ANT-C, which are not defined as homeotic in the strict sense of the word, such as *fushi tarazu* (*ftz*; Wakimoto *et al.*, 1984), *bicoid* (*bcd*; Frohnhöfer and Nüsslein-Volhard, 1985) and *zerknüllt* (*zen*; Wakimoto *et al.*, 1984) are not represented for clarity. Our preliminary work suggests that the *Hox-5* complex may also contain genes with highly divergent or no homeobox sequences, interspersed between homeogenes. As seen in this model, physically linked genes of the mouse (e.g. *Hox-5.2* and *Hox-1.6*) are comparable to *Drosophila* genes that are split between the ANT-C and BX-C complexes (e.g. *AbdB* and *lab*). As a consequence one could postulate that the organism ancestor of both protostomes and deuterostomes already possessed a single continuous cluster of homeobox-containing genes. This original cluster would have been subject to

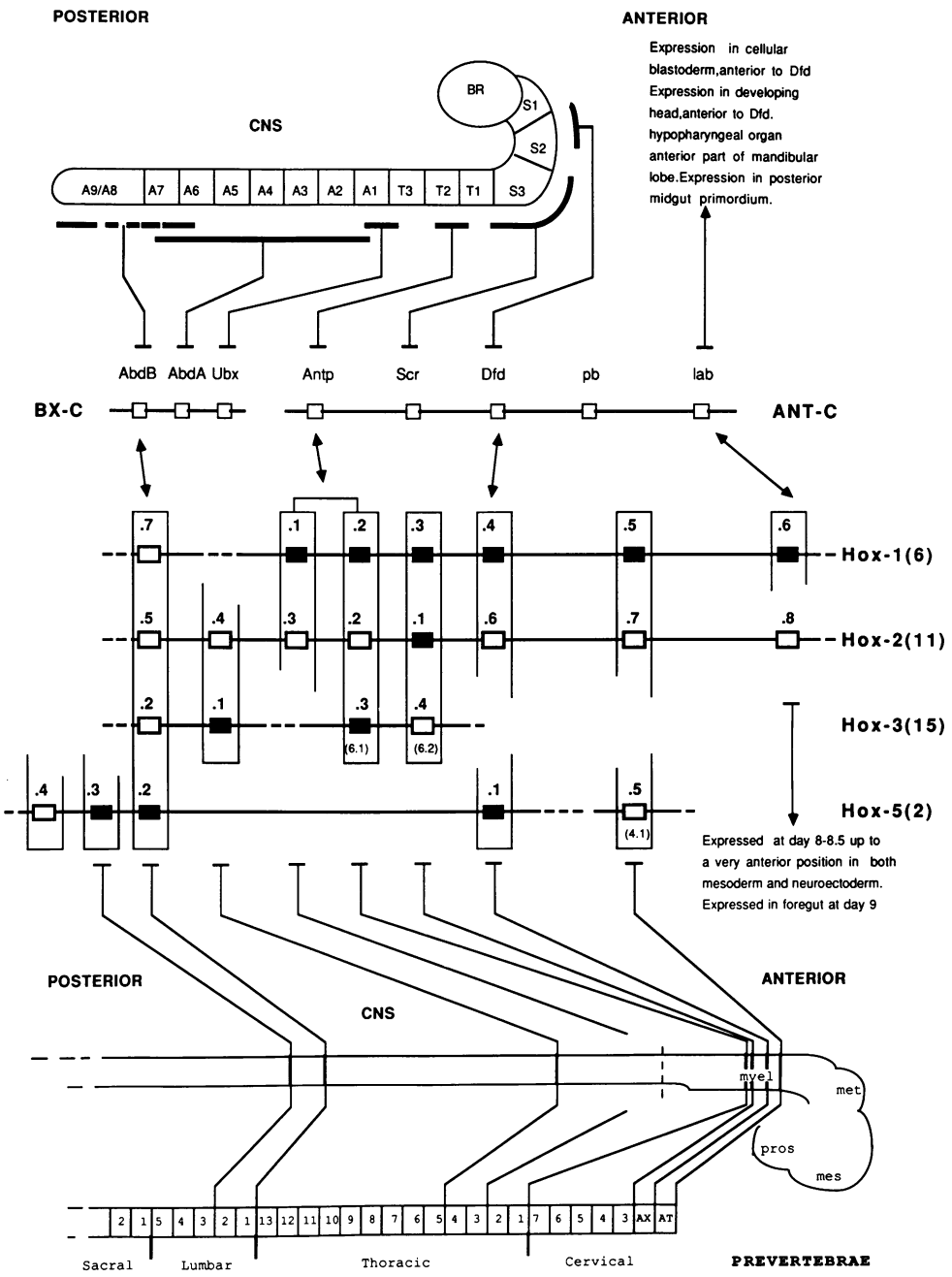

Fig. 6. Schematic representation of the possible correlation between the *Drosophila* homeotic gene complexes and the murine (vertebrate) HOX gene network. The upper part represents the domains of expression of *Drosophila* homeotic genes, members of either BX-C or ANT-C, in the embryonic CNS. The data are essentially taken from Harding *et al.* (1985) and Mlodzick and Gehring (1988); and Hoey *et al.* (1986) for the *labial* (*lab*) gene. Other non-homeotic genes (*zen*, *bic*, *fiz*) located within the ANT-C are not indicated for clarity. The central part represents the HOX complexes, as in Figure 3, with, in filled boxes, genes which have been studied by comparative *in situ* hybridization experiments and whose expression domains (or, at least, the position of their anterior boundaries) have been defined. These boundaries are probably representative of those of all the genes belonging to the same subfamily (indicated by the vertical open or closed rectangles). Thus, genes within the same rectangle are expected to have comparable expression boundaries or, at least, comparable antero-posterior relative positions within their respective complexes (see Gaunt *et al.*, 1988). The bottom part schematically represents the antero-posterior boundaries of expression of these genes along the foetal CNS and pre-vertebral column. In both structures a unique boundary is given for each gene subfamily without considering slight variations which might occur within a particular subfamily (see text). Data are from: *Hox-3.1*, *Hox-1.2*, *Hox-1.4*, *Hox-1.5* (Gaunt *et al.*, 1988), *Hox-1.6* (Sharpe *et al.*, 1988; Gaunt *et al.*, 1988); *Hox-1.1* (K.Mahon and P.Gruss, personal communication); *Hox-5.1* (Featherstone *et al.*, 1988); *Hox-5.2*, *Hox-5.3*, *Hox-1.6* (this work). The upper part is linked to the central and bottom parts with arrows indicating significant amino acid similarities between *Drosophila* and murine genes. Sequence data are from: *AbdB* (Regulski *et al.*, 1987); *Antp* (McGinnis *et al.*, 1984); *Dfd* (Regulski *et al.*, 1985); *lab* (Mlodzick and Gehring, 1988; Hoey *et al.*, 1986), others as in Figure 3. myel, myelencephalon; met, metencephalon; mes, mesencephalon; pro, prosencephalon. AX and AT, axis and atlas, the second and first pre-vertebrae respectively.

duplication events during the evolution of the vertebrates or split in two, in the case of flies. This last event may have occurred more recently since the red flour beetle *Tribolium castaneum* seems to contain a unique cluster of genes comprising homologues of both the ANT-C and BX-C (Beeman, 1987). This latter arthropod might therefore

present a HOX complex reminiscent of that of the common ancestor and the molecular analysis of this complex is certainly of interest in this context. The presence, in an ancestor organism, of such a HOX complex might reflect its coelomate grade of organization. This organism may thus have an organizational level higher than what is generally thought. Alternatively, such a molecular structure may have specified a very simple antero-posterior information in acoelomate triblastic organisms.

The selection for a multi-clustered HOX organization in mammalian species might simply reflect the higher complexity of these organisms and therefore the need for a diversification of the information. Thus each of the members (from one to four) of a given subfamily might display, within the broad antero-posterior specificity of the subgroup, more precise information (for example a sub-specification of the positional information). The comparative analysis of genes belonging to the same subfamily will clarify this point.

At the molecular level, these results suggest that the developmental control mechanisms possibly achieved, both in flies and vertebrates, by the *Hox* gene products might have been conserved. The similarities found at the level of the homeoboxes might reflect the conservation of target binding sites required in a regulatory network (see Desplan *et al.*, 1988; Hoey and Levine, 1988; Müller *et al.*, 1988 and refs therein). In this case, functional aspects may have been conserved while their specific roles may be different. The definition, in vertebrates, of genes regulating or under the regulation of the HOX network will elucidate these questions.

Materials and methods

Cloning of the HOX-5 complex

The pcos2EMBL cosmid library (a gift from H.Lehrach, ICRF, London) was screened with a *Hox-5.1* cDNA clone (Featherstone *et al.*, 1988) and clones covering 60 kd of DNA were isolated. These cosmids were then hybridized to the following homeobox probes under reduced stringency conditions (Duboule *et al.*, 1986): a *Hox-3.1 AvaII–SalI* fragment containing the homeobox (A.Baron and D.Duboule, unpublished results); a *Hox-1.4 EcoRI–BglII* fragment internal to the homeobox (Duboule *et al.*, 1986); a *Hox-1.6 BglII–HpaII* fragment with most of the homeobox (Baron *et al.*, 1987) and two *Hox-1.5* fragments, *BglII–PvuII* and *BglII–EcoRI*, containing parts of the homeobox (McGinnis *et al.*, 1984; Duboule *et al.*, 1986). Two cross-hybridizing *Bam*HI fragments were subcloned in the pPolyIII vector (Lathe *et al.*, 1987) and sequenced using the Maxam and Gilbert method (1977). All filter hybridizations were carried out on Hybond-N sheets (Amersham) with ^{32}P-labelled probes. The restriction map of the cosmids was done using conventional strategies. For *in situ* hybridization, DNA fragments located 3′ from the different homeoboxes (to avoid specific cross-hybridizations to other homeobox-containing sequences) were cloned into the vector pGem-1 (Promega Biotec, Madison, WI). For *Hox-5.2*, the *Bam*HI–*Pvu*II fragment containing the homeobox was cloned and recut with *Fok*I, thus generating the 3′ probe indicated by a shaded rectangle in Figure 1. For *Hox-5.3*, three different probes were tried which gave the same results but with more or less background. The intermediate sized probe proved to be the best and was therefore systematically used. This probe is diagrammed as a shaded rectangle in Figure 1 and is a *Bam*HI–*Hin*dIII fragment extending in the 1.3 kb DNA *Bam*HI fragment 3′ from the *Hox-5.3* subclone enlarged in Figure 1. For *Hox-1.6*, the *BglII–Eco*RI fragment containing the extreme 3′ end of the homeobox plus flanking sequences was used (see Baron *et al.*, 1987, for the restriction map).

Preparation of the RNA probes and in situ hybridizations

^{35}S-Labelled antisense RNA probes (Melton *et al.*, 1984) with a specific activity of ~5 × 10^8 c.p.m./μg were prepared using either T7 (*Hox-5.2*; *Hox-1.6*) or SP6 (*Hox-5.3*) polymerases (Promega), according to the manufacturer's recommendations. The plasmids were previously linearized with *Fok*I, *Hin*dIII and *Eco*RI respectively, and [^{35}S]CTP (850 Ci/mmol, Amersham) was used as substrate. The lengths of the probes were reduced to ~100 nucleotides by limited alkaline hydrolysis with $NaHCO_3$ at pH 10.2 (Cox *et al.*, 1984). The various control sense RNA probes were simultaneously synthesized using the opposite strands as templates. Embryos were obtained from natural matings of laboratory strain animals. Midday of the day of the vaginal plug was designated as day 0.5 p.c. Embryos up to 9.5 days p.c. were fixed, embedded and sectioned in their deciduae, whereas older foetuses were dissected out of their membranes prior to fixation. After fixation in 4% fresh paraformaldehyde in phosphate-buffered saline (12–24 h, 4°C), the embryos were dehydrated in ethanol, cleared in xylene and embedded in paraffin wax (m.p. 56°C). Sections (5–6 μm thick) were collected on 0.5% gelatin/0.5% alum chrome subbed glass microscope slides. The *in situ* hybridization experiments were carried out essentially as described in Gaunt *et al.* (1986) with the following modifications. The slides were prehybridized for 2 h at 50°C, in 50% formamide, 0.3 M NaCl, 10 mM Tris–Cl, pH 6.8, 10 mM $NaPO_4$, pH 6.8, 5 mM EDTA, 1 × Denhardt's, 10 mM DTT, 500 mg/ml yeast RNA, 100 mg/ml salmon sperm DNA and 1.0 mg/ml of non-labelled thio-dUTP (Du Pont). After the RNase A treatment the slides were washed for 1 h in the washing buffer. They were subsequently washed in 2 × SSC for 15 min, at room temperature, then in 0.1 × SSC for 15 min at 50°C, before a final wash in 3 l of 0.1 × SSC for 30 min at room temperature. After dehydration of the sections, they were coated with Kodak NTB 2 emulsion, dried and stored at 4°C. The exposure times were from 12 to 15 days. Kodak D19 developer was used for 2 min at 20°C. The sections were then stained in toluidine blue, dehydrated in ethanol and mounted under coverslips in Eukitt mountant.

Acknowledgements

This work was carried out in the institute of Professor P.Chambon, in Strasbourg. We thank him for his constant support and advice. We thank many colleagues for sharing ideas, unpublished results or technical competences, in particular Robb Krumlauf, Peter Gruss, Steven Gaunt, Walter Gehring, Pierre Gerlinger, Mark Featherstone and Charles Hart. We also thank A.Baron for help in the initial part of this work, B.Schuhbaur for technical assistance, B.Boulay and C.Werle for preparation of figures, the secretarial staff in Strasbourg and H.Davies in Heidelberg for typing the manuscript. This work was supported by grants from INSERM, CNRS, ARC and Fondation pour la Recherche Médicale Française.

References

Akam,M. (1987) *Development*, **101**, 1–22.
Awgulewitsch,A., Utset,M.F., Hart,C.P., McGinnis,W. and Ruddle,F.H. (1986) *Nature*, **320**, 328–335.
Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and Duboule,D. (1987) *EMBO J.*, **6**, 2977–2986.
Beeman,R.W. (1987) *Nature*, **327**, 247–249.
Boncinelli,E., Somma,R., Acampora,D., Pannese,M., D'Esposito,M., Faiella,A. and Simeone,A. (1977) *Hum. Reprod.*, **3**, 880–886.
Breier,G., Dressler,G.R. and Gruss,P. (1988) *EMBO J.*, **7**, 1329–1336.
Carrasco,A.E., McGinnis,W., Gehring,W.J. and DeRobertis,E.M. (1984) *Cell*, **37**, 409–414.
Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 779–789.
Colberg-Poley,A.M., Voss,S.D., Chowdhury,K., Stewart,C.L., Wagner, E.F. and Gruss,P. (1985) *Cell*, **43**, 39–45.
Cox,K.H., DeLeon,D.V., Angerer,L.M. and Angerer,R.C. (1984) *Dev. Biol.*, **101**, 485–502.
Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081–1090.
Dollé,P. and Duboule,D. (1989) *EMBO J.*, **8**, 1507–1515.
Dony,C. and Gruss,P. (1987) *EMBO J.*, **6**, 2965–2975.
Duboule,D., Baron,A., Mähl,P. and Galliot,B. (1986) *EMBO J.*, **5**, 1973–1980.
Duboule,D., Galliot,B., Baron,A. and Featherstone,M.S. (1989) In De Laat,S., Bluemink,J.G and Mummery,C.L. (eds), *Cell to Cell Signals in Mammalian Development*, NATO ASI Series. Springer-Verlag, Heidelberg, pp. 97–108.
Fainsod,A., Awgulewitsch,A. and Ruddle,F.H. (1987) *Dev. Biol.*, **124**, 125–133.
Featherstone,M.S., Baron,A., Gaunt,S.J., Mattei,M.G. and Duboule,D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4760–4764.
Frohnhöfer,H. and Nüsslein-Volhard,C. (1986) *Nature*, **324**, 120–125.
Gaunt,S.J. (1987) *Development*, **101**, 51–60.
Gaunt,S.J. (1988) *Development*, **103**, 135–144.
Gaunt,S.J., Miller,J.R., Powell,D.J. and Duboule,D. (1986) *Nature*, **324**, 662–664.

Gaunt,S.J., Sharpe,P.T. and Duboule,D. (1988) *Development*, **104** (Suppl.), 169–181.
Gehring,W.J. (1987) *Science*, **236**, 1245–1252.
Graham,A., Papalopulu,N., Lorimer,J., McVey,J.-H., Tuddenham,E.G.J. and Krumlauf,R. (1988) *Genes Dev.*, **2**, 1429–1438.
Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, in press.
Harding,K., Weddeen,C., McGinnis,W. and Levine,M. (1985) *Science*, **229**, 1236–1242.
Hart,C.P., Fainsod,A. and Ruddle,F.H. (1987) *Genomics*, **1**, 182–195.
Harvey,R.P., Tabin,C.J. and Melton,D.A. (1986) *EMBO J.*, **5**, 1237–1244.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
Hoey,T., Doyle,H.J., Harding,K., Wedeen,C. and Levine,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4809–4813.
Holland,P.W.H. and Hogan,B.L.M. (1986) *Nature*, **321**, 252–253.
Holland,P.W.H. and Hogan,B.L.M. (1988a) *Genes Dev.*, **2**, 773–782.
Holland,P.W.H. and Hogan,B.L.M. (1988b) *Development*, **102**, 159–174.
Kaufmann,T.C., Lewis,R. and Wakimoto,B.T. (1980) *Genetics*, **94**, 115–133.
Lathe,R., Vilotte,J.L. and Clark,A.J. (1987) *Gene*, **57**, 193–201.
LeMouellic,H., Condamine,H. and Brulet,P. (1988) *Genes Dev.*, **2**, 125–135.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lonai,P., Arman,E., Czosnek,H., Ruddle,F.H. and Blatt,C. (1987) *DNA*, **6**, 409–418.
Martinez-Arias,A., Ingham,P.W., Scott,M.P. and Akam,M.E. (1987) *Development*, **100**, 573–683.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
McGinnis,W. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 263–270.
McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Nature*, **308**, 428–433.
Melton,D., Krieg,P., Rebagliati,M., Maniatis,T., Zinn,R. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035.
Mlodzick,M., Fjose,A. and Gehring,W.J. (1988) *EMBO J.*, **7**, 2569–2578.
Müller,M., Affolter,M., Leupin,W., Otting,G., Wühtrich,K. and Gehring,W.J. (1988) *EMBO J.*, **8**, 4299–4304.
Oliver,G., Wright,C.V.E., Hardwicke,J. and De Robertis,E.M. (1988) *EMBO J.*, **7**, 3199–3209.
Otting,G., Qian,Y.Q., Müller,M., Affolter,M., Gehring,W.J. and Wühtrich,K. (1988) *EMBO J.*, **8**, 4305–4309.
Pardue,M.L. (1985) In Hames,B.D. and Higgins,S.J. (eds), *Nucleic Acid Hybridization. A Practical Approach*. IRL Press, Oxford, pp.
Pultz,M.A., Diederich,R.J., Cribbs,D.L. and Kaufmann,T.C. (1988) *Genes Dev.*, **2**, 901–920.
Regulski,M., Harding,K., Kostriken,R., Karch,F., Levine,N. and McGinnis,W. (1985) *Cell*, **43**, 71–80.
Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 767–777.
Rubin,M.R., King,W., Toth,L.E., Sawczuk,I.S., Levine,M.S., D'Eustachio,P. and Nguyen-Huu,M.C. (1987) *Mol. Cell. Biol.*, **7**, 3836–3841.
Sanchez-Herrero,E., Vernos,I., Marco,R. and Morata,G. (1985) *Nature*, **313**, 108–113.
Scott,M.P. and Carroll,S.B. (1987) *Cell*, **51**, 689–698.
Scott,M.P., Tamkun,I.W. and Harzell,G.W., III (1988) *BBA Reviews on Cancer*, in press.
Sharpe,P.T., Fritz,A., De Robertis,E.M. and Gurdon,J.B. (1987) *Cell*, **50**, 749–758.
Sharpe,P.T., Miller,J.R., Evans,E.P., Burtenshaw,M.D. and Gaunt,S.J. (1988) *Development*, **102**, 397–407.
Simeone,A., Pannese,M., Acampora,D., D'Esposito,M. and Boncinelli,E. (1988) *Nucleic Acids Res.*, **16**, 6379.
Theiler,K. (1972) *The House Mouse*. Springer-Verlag, Berlin.
Utset,M.F., Awgulewitsch,A., Ruddle,F.H. and McGinnis,W. (1987) *Science*, **235**, 1379–1382.
Wakimoto,B.T., Turner,F.R. and Kaufmann,T.C. (1984) *Dev. Biol.*, **102**, 147–172.

Received on January 9, 1989; revised on February 17, 1989

Note added in proof

Hox5.2 and *Hox5.3* sequences have received the accession numbers X14714 and X14715, respectively, in the EMBL data bank.